

Automatic pathology classification using a single feature machine learning - support vector machines

Fernando Yepes-Calderon^{b,c}, Fabian Pedregosa^e, Bertrand Thirion^e, Yalin Wang^{d,*} and
Natasha Leporé^{a,b,*}

^aUniversity of Southern California, 900 W 34th St, Los Angeles-CA, USA;

^bChildren’s Hospital Los Angeles, 4650 Sunset Blvd, Los Angeles-CA, USA;

^cUniversidad de Barcelona, Carrer de casanova 143, Barcelona-Spain;

^dComputer Science and Engineering, Arizona State University, Tempe, AZ, USA;

^eParietal Team INRIA Saclay-Ile-de-France, France

ABSTRACT

Magnetic Resonance Imaging (MRI) has been gaining popularity in the clinic in recent years as a safe in-vivo imaging technique. As a result, large troves of data are being gathered and stored daily that may be used as clinical training sets in hospitals. While numerous machine learning (ML) algorithms have been implemented for Alzheimer’s disease classification, their outputs are usually difficult to interpret in the clinical setting. Here, we propose a simple method of rapid diagnostic classification for the clinic using Support Vector Machines (SVM)¹ and easy to obtain geometrical measurements that, together with a cortical and sub-cortical brain parcellation, create a robust framework capable of automatic diagnosis with high accuracy. On a significantly large imaging dataset consisting of over 800 subjects taken from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database, classification-success indexes of up to 99.2% are reached with a single measurement.

Keywords: Alzheimer’s disease, machine learning, mild cognitive impairment, support vector machines, fast clinical diagnosis.

1. INTRODUCTION

Magnetic Resonance Imaging is highly popular, and its growing use in the clinic is producing vast amounts of data. This stored data is often difficult to interpret visually, and, after the initial reading, does not currently provide added clinical value other than as a retrospective reference for patients and clinicians. However, from the perspective of machine learning, it could be the keystone for accelerating the diagnosis of diseases. Moreover, machine learning methods can be applied to generate accurate prognosis, provided that a previous classification stage is accomplished with rigorousness. If a good experimental setup is accomplished, the accuracy of a machine learning implementation increases with the number of independent features that it can learn from and the number of training sets that it is fed. Conveniently, hospitals can provide a large source of data as training sets.

Here we present a machine learning - support vector machines design that is able to classify subjects into Alzheimer’s disease, mild cognitive impairment and controls by limiting itself to four individual geometrical measurements among 35 cortical regions in the brain. Our formulation is motivated by the real clinical scenario, where the measurements have to be simple, intuitive and explainable. When a new subject is scanned, our system labels it in one of three groups: Alzheimer’s disease, mild cognitive impairment or control, using previous knowledge that exists intrinsically in the clinical database.

* equal senior author contribution

2. MATERIALS AND METHODS

Our dataset consists of 829 patients from the ADNI (adni.loni.usc.edu) database where subjects are classified into three groups by well-established clinical diagnosis criteria: Alzheimer’s disease (AD), mild cognitive impairment (MCI) and Controls (CTL) with 196, 406 and 228 subjects per group respectively. Using FreeSurfer,^{2,3} measurements of structural thickness, contour area, volume and structural curvature are extracted in 35 regions of the brain. For implementation purposes, these regions are labeled from 1 to 35.

The data is presented in comma separated values (CSV) files that are integrated in a Python numpy matrix. This platform was chosen intentionally to comply with the Scikit-learn machine learning framework. The numpy matrix is later queried using Python list comprehensions for data filtering, segmentation, scaling and normalization. The classifiers’ estimators are proved by cross correlation. The training and test samples are gathered from the original dataset with a ratio of 0.8 : 0.2 (train:test) in a 6-fold experimental scheme. The scores and the records of sensitivity and specificity are the mean values of the 6 experiments. With this data, three different setups are studied (see Sections 2.2, 2.3 and 2.4).

2.1 Exploring data patterns through principal component analysis and preliminary experiments

Despite the fact that ML implementation proposed here has low dimensionality when compared with common uses of ML algorithms, the 140 dimensions (see Section 2.4) are difficult to represent for direct inspection and classification. Instead, we use a principal component analysis (PCA) to project points from a high n-dimensional field to a lower m-dimensional one. In our case, $m = 2$ is used for plotting purposes.

The PCA plots (see Figure 1), suggested a high degree of class separability when comparing AD with CTL patients, while for the other possible groupings the distributions are more complex. We looked for quantitative indexes of classification through exploratory implementations of SVM using a linear kernel. These experiments and their setups are explained below.

2.2 Approach A: mean value per region

The first experimental approach consists of using the mean value per region of each geometrical factor as a feature to see the influence of the geometrical factor in the classification process and discover regions of high differentiation. This setup results in a study matrix of n rows by 1 column. The experiment is run 12 times and is the product of the four geometrical factors and the three pair of diagnosis groups among the pathologies. The n values for these experiments are shown in the Section 3 on top of Tables 2 and 2.

2.3 Approach B: histogram

Here, we face a trade-off between accuracy and applicability. For high accuracy, it is possible to try a voxel-based approach, but this requires time consuming computations that may reduce its value for the clinic. Alternatively, in the last section, we computed a mean value of the geometrical factor per brain region, thus ML is easily implementable; nevertheless the averaging function may smooth out some important differences. In order to limit computational time while avoiding an excessive detail reduction, we decided to use a histogram approach and create feature bins for our SVM implementation.

The dataset matrix results in n samples x 35 columns x number of bins. the 35 factor corresponds to the number of regions studied in the brain, which is multiplied by the number of bins in the histogram. Here, we limit ourselves to experiments for which we divide the data into 2 bins. The experiment above is repeated 12 times to test all the geometrical factors. The number of samples per diagnosis group (n), is given in Section 3 on top of tables 3 and 4.

2.4 Approach C: feature reduction and pathology linked brain regions

Improvement of the classification indexes was done through dimension reduction using a K-best feature extraction. In addition to providing the best classifying geometrical factors, this method yields the cerebral regions that are most useful for classification.

The brain regions are selected by a feature reduction algorithm, that iteratively calculates the f value while including region-by-region and factor-by-factor analyses. After each iteration, a SVM estimator is calculated and a cross correlation experiment is run. The iteration that provides the highest score is selected along with the regions and geometrical factors that best discriminate between the diagnoses. This setup results in a matrix of n subjects x 35 brain regions x 4 geometric factors. The value n is the same one that was used in experiments explained in Section 2.3 for each grouping test.

3. RESULTS

The high dimensionality reached when the four geometrical factors are combined with the 35 brain regions analyzed make data trends difficult to visualize. Instead, we use PCA to reduce the dimension of the dataset, see Figure 1.

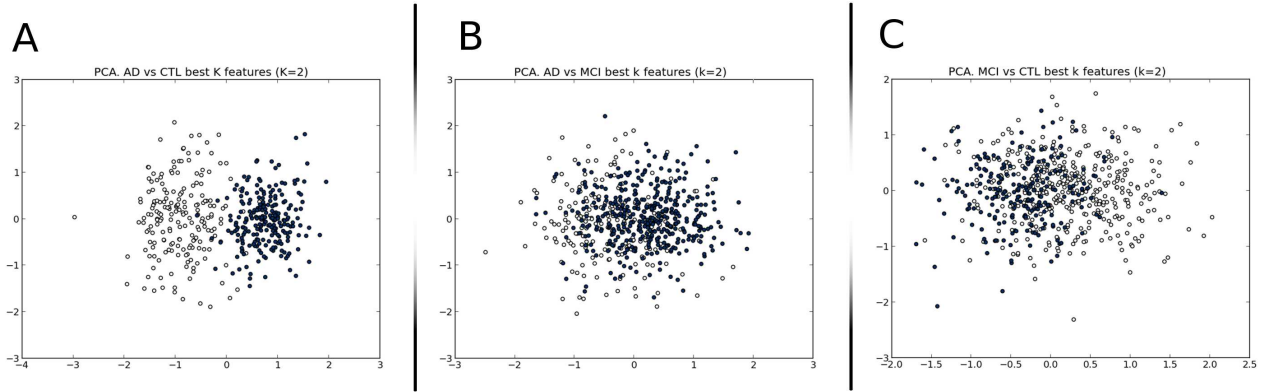


Figure 1: Principal Component Analysis. Frame A. AD vs. CTL. Frame B. AD vs. MCI. Frame C. MCI vs. CTL.

Figure 1, panel A shows that a classifying algorithm using linear conditioning may create a separating hyperplane which will accurately classify new subjects between the AD and CTL groups. As for the other two groups, no linear boundaries are detected.

The resulting scores of an out-of-the-box linear SVM operating on the data including all the regions in all subjects as samples and each geometrical factors (one-by-one) as features are shown in Tables 1 and 2.

AD vs. CTL (14840 samples)				AD vs. MCI (21070 samples)				MCI vs. CTL (22190 samples)			
Th.	Area.	Vol.	Curv.	Th.	Area.	Vol.	Curv.	Th.	Area.	Vol.	Curv.
57.3	55.1	55.5	53.9	67.4	67.2	67.5	67.9	64.0	64.3	64.0	63.8
s: 44.2	s: 68.6	s: 45.4	s: 51.5	s: 66.2	s: 70.6	s: 72.0	s: 55.9	s: 60.8	s: 52.7	s: 42.4	s: 65.6
e: 60.1	e: 61.5	e: 54.7	e: 50.1	e: 55.8	e: 65.6	e: 66.3	e: 58.1	e: 66.8	e: 59.2	e: 59.9	e: 56.7

Table 1: SVM scores in the left hemisphere (LH) for the mean value per region (see Section 2.2)

AD vs. CTL (14840 samples)				AD vs. MCI (21070 samples)				MCI vs. CTL (22190 samples)			
Th.	Area.	Vol.	Curv.	Th.	Area.	Vol.	Curv.	Th.	Area.	Vol.	Curv.
85.6	68.6	84.4	62.7	71.7	69.1	72.4	67.4	68.4	62.3	68.7	63.8
s: 92.4	s: 66.9	s: 91.4	s: 61.2	s: 78.3	s: 72.5	s: 80.9	s: 66.6	s: 72.7	s: 65.1	s: 69.1	s: 59.3
e: 81.7	e: 55.5	e: 82.3	e: 67.7	e: 71.4	e: 67.7	e: 72.6	e: 45.1	e: 71.1	e: 60.0	e: 55.9	e: 65.5

Table 2: **SVM scores in the right hemisphere (RH) for the mean value per region (see Section 2.2)**

The mean computation described in Section 2.2 yields low accuracy; the sensitivity (s) and specificity (e) are low as well. We then look at the histograms as described in Section 2.3. The results of a 2-bins histogram setup are shown in Tables 3 and 4.

AD vs. CTL (162 samples)				AD vs. MCI (301 samples)				MCI vs. CTL (317 samples)			
Th.	Area.	Vol.	Curv.	Th.	Area.	Vol.	Curv.	Th.	Area.	Vol.	Curv.
82.7	66.8	83.9	53.9	78.2	67.5	71.5	67.5	69.4	64.0	66.0	64.0
s: 82.8	s: 66.6	s: 81.8	s: 56.9	s: 77.4	s: 72.4	s: 75.0	s: 78.8	s: 73.8	s: 64.7	s: 75.0	s: 73.8
e: 88.3	e: 69.5	e: 83.6	e: 59.5	e: 75.9	e: 67.8	e: 71.6	e: 75.0	e: 58.9	e: 40.9	e: 52.9	e: 58.9

Table 3: **SVM scores in the left hemisphere (LH) for the 2-bins histograms**

AD vs. CTL (162 samples)				AD vs. MCI (301 samples)				MCI vs. CTL (317 samples)			
Th.	Area.	Vol.	Curv.	Th.	Area.	Vol.	Curv.	Th.	Area.	Vol.	Curv.
85.6	68.6	84.4	62.7	71.7	69.1	72.4	67.4	68.4	62.3	68.7	63.8
s: 92.4	s: 66.9	s: 91.4	s: 61.2	s: 78.3	s: 72.5	s: 80.9	s: 66.6	s: 72.7	s: 65.1	s: 69.1	s: 59.3
e: 81.7	e: 55.5	e: 82.3	e: 67.7	e: 71.4	e: 67.7	e: 72.6	e: 45.1	e: 71.1	e: 60.0	e: 55.9	e: 65.5

Table 4: **SVM scores in the right hemisphere (RH) for the 2-bins histograms**

For tables 1, 2, 3 and 4, Th stands for thickness, Vol for volume and Curv for curvature. The scores are shown in percentile (%) values and represent the index of success of the algorithm at classifying the subjects.

While the indexes reported above are all over the guessing level, they are still too low for usability in the clinic. Since neurological disorders do not affect the brain uniformly, we design an experiment where a feature reduction operation can discard brain regions while increasing the score of the estimator. Figure 2 depicts the selected regions while Table 5 shows the scores obtained when the SVM algorithm works only in the reduced dataset.

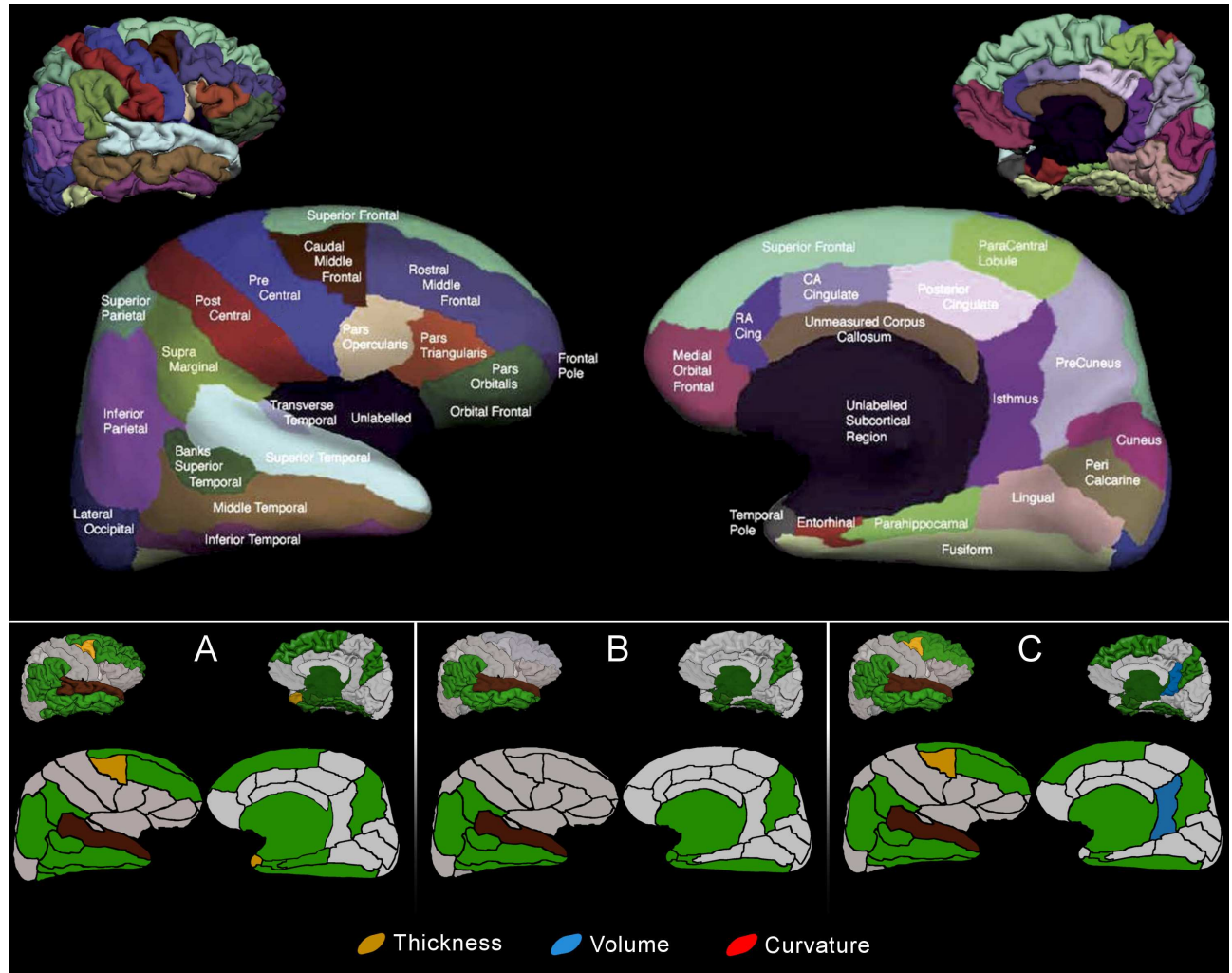


Figure 2: Cortical and sub-cortical brain regions with high impact when using SVM and geometrical factors for automatic diagnosis. Frame A. AD vs CTL. Frame B. AD vs MCI. Frame C. MCI vs CTL. The top panel is an adaptation of the Desikan Atlas⁴ used here for naming and positioning reference. The legend use primary colors intentionally so the reader can infer in which regions a geometrical factor is relevant in stand alone fashion or by combining it with another factor p.e: green refers to regions where blue and yellow coded factors act together, brown defines a region where the three geometrical factors are relevant.

	AD vs. CTL (162 Samples)	AD vs. MCI (301 Samples)	MCI vs. CTL (317 Samples)
LH	99.2 — s: 98.9 e: 97.8	86.3 — s: 90.1 e: 85.6	90.2 — s: 92.4 e: 89.6
RH	98.7 — s: 97.3 e: 98.0	88.3 — s: 92.4 e: 89.0	91.1 — s: 93.4 e: 90.4

Table 5: SVM scores in both hemispheres after applying the k-best feature reduction function

The values in Table 5 must be complemented with findings depicted in Figure 2. A well driven interpretation must go as follows. The score in group AD vs. CTL in the left hemisphere is 98.7% when the colored regions are used as factors together with the geometrical measurements that yields the color code used in the panel A of Figure 2. Recall that the score is the result of averaging a 6-fold cross correlation experiment. A similar

reasoning must be used to interpret the scores in the other diagnostic pairs.

4. DISCUSSION

Machine learning (ML) is an outstanding tool for detecting patterns in high dimensional data sets such as the one used here. The goal of this work was to provide a simple and fast detection tool to classify controls, MCI and AD in the clinic. To do so, we used the well established Scikit-learn software, which avoids common ML problems such as biasing or data variance. In the real clinical environment, a fast but accurate imaging measurement is most useful as a first step toward additional tests. So we restricted our results to some reliable but simple imaging measurement as the features for the classification. For all trials, the combination of thickness and volume provided the two best discriminating measures.

From a purely numerical point-of-view, it is worth noting that the area is not included in Figure 2. This is due to the fact that the area resulted in low sensitivity and was always selected out in the feature reduction stage. With the remaining three factors and the 35 brain regions, we were able to automatically distinguish between a healthy brain and one suffering from AD with a success index reaching 98.7% and 99.2% for the right and left hemispheres respectively. This index is above the 85.7% index reported in⁵ and 82.8% in⁶ for the same classification. When trying to differentiate between AD and MCI brains, our best index reached 88.3%. This classification test was not done in,⁵ nor in.⁶ However,⁵ reported high accuracy in predicting MCI converters to AD (86%), something that we have not explored yet with our approach. The authors in⁵ proposed a complex weighted scheme in the whole brain that, in addition to the geometrical measurements, includes social factors as features, something difficult to assess in the clinic.

The results presented in this work are to be also compared with other techniques to distinguish between diagnosis groups. With the increasing use of the MRI, many authors have proposed deterministic methods to rapidly diagnose AD patients and predict MCI converters. Some of these references include,⁷ which used coronal T1-weighted visual rating scales and yielded a 80 – 85% sensitivity and specificity to distinguish patients with AD from those with no cognitive impairment (AD vs CTL group in this manuscript), and only slightly lower sensitivity and specificity levels for diagnosing amnesic MCI. These scales also had good predictive power to anticipate decline in MCI. Other authors have focused on specific brain regions and have associated physiological facts that might help to increase the accuracy and the speed of automatic diagnosis tools. According to⁸ at the mild dementia stage of AD, hippocampal volume is already reduced by 15 – 30% relative to controls, and in the amnesic variant of MCI the volume is reduced by 10 – 15%. A recent meta-analysis estimated that medial temporal atrophy has 73% sensitivity and 81% specificity for predicting whether patients with amnesic MCI will convert to dementia.⁹ If medial temporal atrophy is measured with a continuous metric such as hippocampal volume, specificity might be increased, but at the cost of reduced sensitivity.¹⁰

5. CONCLUSIONS

For any diagnostic marker to be useful in practice, its capacity to separate two or more conditions that can be confused on clinical grounds is a necessary but not sufficient requirement. Clinical usefulness requires that the marker provides benefit over and above that provided by clinical assessment.¹⁰ Here, we present a method that accomplish this. Moreover, the MRI acquisition system and the software used to recover the data here are useable in research and in clinical settings, hence facilitating translation of our method from the former to the later. Gathering information and interpreting it through machine learning methods opens not only the opportunity to use sophisticated MRI analyses methods in the medical field, but also to make health care institutions more time effective, hence improving patient care.

REFERENCES

- [1] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

- [2] Fischl, B. and Dale, A., "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *Proceedings of the National Academy of Sciences of the United States of America* **97**(20), 11050–11055 (2000).
- [3] Han, X., Jovicich, J., Salat, D., Fennell-Clay, V., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., K., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B., and Fischl, B., "Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer," *NeuroImage* **32**(1), 180–194 (2006).
- [4] Desikan, R., Segonne, A., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, P., Hyman, B., Albert, M., and Killiany, R., "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *NeuroImage* **31**(1), 968–80 (2006).
- [5] Aguilar, C., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Spenger, C., Simmons, A., and Wahlund, L.-O., "Different multivariate techniques for automated classification of mri data in alzheimer's disease and mild cognitive impairment,"
- [6] Cuingnet, R. et al., "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage* **56**, 766 – 781 (15 May 2011).
- [7] Scheltens, P., Leys, D., Barkhof, F., Huglo, D., Weinstein, H., Vermersch, P., Kuiper, M., Steinling, M., Wolters, E. C., and Valk, J., "Atrophy of medial temporal lobes on mri in "probable" alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates.," *Journal of Neurology, Neurosurgery & Psychiatry* **55**(10), 967–972 (1992).
- [8] Shi, F., Liu, B., Zhou, Y., Yu, C., and Jiang, T., "Hippocampal volume and asymmetry in mild cognitive impairment and alzheimer's disease: Meta-analyses of mri studies," *Hippocampus* **19**(11), 1055–1064 (2009).
- [9] Yuan, Y., Gu, Z.-X., and Wei, W.-S., "Fluorodeoxyglucose-positron-emission tomography, single-photon emission tomography, and structural mr imaging for prediction of rapid conversion to alzheimer disease in patients with mild cognitive impairment: a meta-analysis," *American Journal of Neuroradiology* **30**(2), 404–410 (2009).
- [10] Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., and Thompson, P. M., "The clinical use of structural mri in alzheimer disease," *Nature Reviews Neurology* **6**(2), 67–77 (2010).